SFB 649 Discussion Paper 2012-049

Simultaneous test procedures in terms of p-value copulae

Thorsten Dickhaus*
Jakob Gierl*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Simultaneous test procedures in terms of *p*-value copulae*

Thorsten Dickhaus and Jakob Gierl
Department of Mathematics
Humboldt-University
Berlin, Germany
Email: dickhaus@math.hu-berlin.de

***Abstract*—At least since [1], a broad class of multiple comparison procedures, so-called simultaneous test procedures (STPs), is established in the statistical literature. Elements of an STP are a testing family, consisting of a set of null hypotheses and corresponding test statistics, and a common critical constant. The latter threshold with which each of the test statistics has to be compared is calculated under the (joint) intersection hypothesis of all nulls. Under certain structural assumptions, the so-constructed STP provides strong control of the family-wise error rate. More recently, a general method to construct STPs in the case of asymptotic (joint) normality of the family of test statistics has been developed in [2], and numerical solutions to compute the critical constant in such cases were provided.**

Here, we propose to look at the problem from a different perspective. We will show that the threshold can equivalently be expressed by a quantile of the copula of the family of *p*-values associated with the test statistics, assuming that each of these *p*-values is marginally uniformly distributed on the unit interval under the corresponding null hypothesis. This offers the opportunity to exploit the rich and growing literature on copula-based modeling of multivariate dependency structures for multiple testing problems and in particular for the construction of STPs in non-Gaussian situations.

JEL Classification: C12, C44.
MSC 2010 subject classifications: 62J15, 62F03, 60E05.

***Keywords*-distributional transform; family-wise error rate; multiple hypotheses testing; multiplicity correction; simultaneous statistical inference; single-step test**

I. Introduction

In many modern areas of science, several inferential problems have to be solved simultaneously on the basis of only one single dataset. Consider, for instance, gene expression analyses where many genetic loci are tested in parallel for differential expression between groups of individuals. Mathematically, such problems can be formalized as multiple statistical hypotheses test problems and the data-analytic tools to solve them are multiple test procedures (MTPs). The field of multiple hypotheses testing has become one of the major branches of mathematical and applied statistics during the past approximately 20 years, especially driven by the need for new data-analytic tools for problems from modern life sciences, such as genetics, proteomics, functional magnetic resonance imaging, or brain-computer interfacing. Although the literature on MTPs is exponentially increasing over time, it is still possible to systematize the proposed methods according to some general categories: For instance, one class of methods only models the marginal distributions of the involved test statistics explicitly and combines these test statistics or, equivalently, corresponding p-values following probabilistic calculations. Examples of this kind of procedures are single-step multiple tests like the classical Bonferroni and Šidák corrections, step-up multiple tests like the famous false discovery rate-controlling linear step-up test by Benjamini and Hochberg (see [3]), step-down tests like Holm's procedure (see [4]), or general step-up-down tests as introduced in [5]. Another class of MTPs considers the full joint distribution of all test statistics and relies on calculating or approximating quantiles of this joint distribution, for instance by resampling (cf. [6], [7]) or by proving asymptotic normality by means of central limit theorems (see, for example, [2] and [8]).

In the present work, we contribute to the theory of the latter class of MTPs. However, we propose only to infer the dependency structure of the involved test statistics, because their marginal distributions are often already implied by the statistical model. Assuming these marginals as fixed, the problem reduces to considering the copula function of the test statistics or the p-values. In Section II, we express simultaneous test procedures (STPs) in terms of the copula function of p-values. STPs take a quantile of the joint distribution of test statistics under the global null hypothesis (all hypotheses are assumed to hold true) as threshold. The section concludes with a general construction principle for STPs based on copulae. In Section III, applications of this general theory are discussed. We will elucidate that our copula view toward STPs is useful for importance weighting of the hypotheses and for separating the marginal models from a model for the dependency structure. To the best of our knowledge, the modeling approach in the present work is novel to the field of multiple testing. The possibility to employ copula-based models for constructing multiple tests has been mentioned in [9], but we are not aware of concrete references realizing this suggestion. The usage of copulae as model diagnosis tools in the context of multiple testing

*This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

is exemplified in [10].

II. General Theory

Throughout the work, we let $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ denote a statistical model. We identify hypotheses with non-empty subsets of the parameter space Θ. The tuple $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H}_m)$ denotes a multiple test problem, where $\mathcal{H}_m = (H_i, 1 \leq i \leq m)$ defines a finite family of m null hypotheses. The resulting alternative hypotheses are denoted by $K_i = \Theta \setminus H_i$, $1 \leq i \leq m$. The intersection hypothesis $H_0 = \bigcap_{i=1}^m H_i$ will occasionally be referred to as global hypothesis. Throughout the work, we assume that H_0 is non-empty. For a given $\vartheta \in \Theta$, we denote the index set of true null hypotheses in $\mathcal{H}_m$ by $I_0 \equiv I_0(\vartheta) = \{1 \leq i \leq m : \vartheta \in H_i\}$. A (non-randomized) multiple test is a measurable mapping $\varphi = (\varphi_i)_{1\leq i\leq m} : \Omega \to \{0,1\}^m$ the components of which have the usual interpretation of a statistical test for H_i versus K_i. The family-wise error rate, FWER for short, of a multiple test φ is (for a given $\vartheta \in \Theta$) defined as the probability under ϑ of at least one type I error, i. e., $\text{FWER}_\vartheta(\varphi) = \mathbb{P}_\vartheta\left(\bigcup_{i\in I_0(\vartheta)}\{\varphi_i = 1\}\right)$ and φ is said to control the FWER at a pre-specified level $\alpha \in (0,1)$ if $\sup_{\vartheta\in\Theta} \text{FWER}_\vartheta(\varphi) \leq \alpha$.

In this work, we restrict our attention to a special class of multiple tests, namely, simultaneous test procedures as originally defined in [1].

Definition 1: Consider the extended test problem $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H}_{m+1})$ with $\mathcal{H}_{m+1} = \{H_i, i \in I^* := \{0, 1, \ldots, m\}\}$. Assume real-valued test statistics $T_i, i \in I^*$, which tend to larger values under alternatives. Then we call

(a) $(\mathcal{H}_{m+1}, \mathcal{T})$ with $\mathcal{T} = \{T_i, i \in I^*\}$ a testing family.
(b) $\varphi = (\varphi_i, i \in I^*)$ a simultaneous test procedure, if

$$\forall 0 \leq i \leq m : \varphi_i = \begin{cases} 1, & \text{if} \quad T_i > c_\alpha, \\ 0, & \text{if} \quad T_i \leq c_\alpha, \end{cases}$$

where the critical value c_α is determined such that $\forall \vartheta \in H_0 : \mathbb{P}_\vartheta(\{\varphi_0 = 1\}) = \mathbb{P}_\vartheta(\{T_0 > c_\alpha\}) \leq \alpha$.

In order to simplify the definition of p-values under our setup, we now impose three important structural assumptions on the test problem and the testing family.

(S1) Any $\vartheta \in H_0$ is a least favorable parameter configuration (LFC) for the FWER of any STP φ for $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H}_m)$ that is based on $(T_1, \ldots, T_m)$, meaning that the FWER of φ becomes largest under parameters in the global hypothesis.
(S2) Every null hypothesis H_i, $1 \leq i \leq m$, is of the form $H_i : \{\theta_i(\vartheta) = \theta_i^*\}$, where $\theta : \Theta \to \Theta'$ denotes a derived parameter, i indexes components of θ, and the θ_i^* are fixed given values in Θ'.
(S3) The marginal cumulative distribution function (cdf.) of T_i under H_i, which we will denote by F_i, is continuous and strictly increasing.

Heuristically, assumption (S1) seems highly plausible: The more true null hypotheses in $\mathcal{H}_m$, the larger the probability of one or more type I errors. If all marginal tests are stochastically independent and once c_α is fixed, the investigations in [11] and [12] show that this reasoning is indeed true. For models with dependency, however, the determination of LFCs becomes much more complicated. In the special case that θ is the identity mapping, sufficient conditions for the validity of (S1) for single-step multiple tests (including STPs) have been derived in [8]. We will discuss examples under our more general setup in the subsequent sections.

Under (S2) - (S3), appropriate p-values corresponding to the T_i are given by $\forall 1 \leq i \leq m : p_i = 1 - F_i(T_i)$. This transformation with the upper-tail cdf. is useful for multiple testing, because it standardizes all marginal test statistics. Every p_i is supported on the unit interval $[0,1]$, even if the T_i have drastically different scales.

The following obvious lemma summarizes further properties of the p_i, $1 \leq i \leq m$.

Lemma 1: Under (S2) - (S3), it holds:

(a) $T_i > c_\alpha \iff p_i < 1 - F_i(c_\alpha)$. We may think of $\alpha_{\text{loc.}}^{(i)} := 1 - F_i(c_\alpha)$ as a multiplicity-adjusted local significance level.
(b) $1 - p_i$ is equal to the distributional transform of T_i as defined in [13].
(c) Under H_i, we have $p_i \sim \text{UNI}[0,1]$ and $1 - p_i \sim \text{UNI}[0,1]$.

The usefulness of p-values for formulating STPs is mainly based on the following well-known theorem.

Theorem 1 (Sklar, cf. [14], [15]):
Let $X = (X_1, \ldots, X_m)^\top$ a random vector with values in $\mathbb{R}^m$ and with joint cdf F_X and marginal cdfs $F_{X_1}, \ldots, F_{X_m}$. Then there exists a function $C : [0,1]^m \to [0,1]$ such that for all $x = (x_1, \ldots, x_m)^\top \in \bar{\mathbb{R}}^m$, it holds

$$F_X(x) = C(F_{X_1}(x_1), \ldots, F_{X_m}(x_m)).$$

If all m marginal cdfs are continuous, the copula C is unique.

From Theorem 1 and Lemma 1.(c) we conclude that, under any $\vartheta^* \in H_0$, the joint cdf. of $(1 - p_i : 1 \leq i \leq m)$ coincides with their copula. We are now ready to bound the FWER of any STP for $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H}_m)$ in terms of the copula of the distributional transforms.

Theorem 2: Let φ an STP for $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H}_m)$. For arbitrary $\vartheta \in \Theta$ and $\vartheta^* \in H_0$, we get under (S1) - (S3) that

$$\text{FWER}_\vartheta(\varphi) \leq 1 - C_{\vartheta^*}(1 - \alpha_{\text{loc.}}^{(1)}, \ldots, 1 - \alpha_{\text{loc.}}^{(m)}), \qquad (1)$$

with C_{ϑ^*} denoting the copula of $(1 - p_i : 1 \leq i \leq m)$ under ϑ^*.

Proof: Due to Lemma 1.(a), it holds

$$\text{FWER}_\vartheta(\varphi) = \mathbb{P}_\vartheta\left(\bigcup_{i\in I_0(\vartheta)}\{p_i < \alpha_{\text{loc.}}^{(i)}\}\right). \qquad (2)$$

Making use of assumption (S1) and the fact that all null hypotheses are true under ϑ^*, we can bound the right-hand side of (2) from above and obtain

$$\begin{aligned}\text{FWER}_\vartheta(\varphi) &\leq \mathbb{P}_{\vartheta^*}\left(\bigcup_{i=1}^m \{p_i < \alpha_{\text{loc.}}^{(i)}\}\right) \\ &= 1 - \mathbb{P}_{\vartheta^*}\left(\bigcap_{i=1}^m \{1-p_i \leq 1-\alpha_{\text{loc.}}^{(i)}\}\right) \\ &= 1 - C_{\vartheta^*}(1-\alpha_{\text{loc.}}^{(1)}, \ldots, 1-\alpha_{\text{loc.}}^{(m)}),\end{aligned}$$

where we used Theorem 1 and Lemma 1.(c) in the last line. ∎

To control the FWER at level α with the STP φ, we can therefore equivalently compare the (marginal) distributional transforms with a suitable $(1-\alpha)$-quantile of their copula under ϑ^*. In particular, the STP re-formulation in (1) does not require an explicit test statistic T_0 for testing H_0.

III. Applications

A. Illustrative example: one-factorial analysis of variance (ANOVA1)

Definition 2 (Dunnett contrasts under ANOVA1): Fix an integer k (number of treatment groups) and sample sizes $(n_i)_{1\leq i\leq k}$, and model the observation $x \in \Omega = \mathbb{R}^{\sum_{i=1}^k n_i}$ as a realization of $X = (X_{i,j} : 1 \leq i \leq k, 1 \leq j \leq n_i)$. In this, assume that

(i) all $X_{i,j}$ are stochastically independent,
(ii) $X_{i,j} \sim \mathcal{N}(\mu_i, 1)$ (or with unknown, but common variance).

The parameter of this model is the unknown mean vector $\mu = (\mu_1, \ldots, \mu_k)^\top \in \mathbb{R}^k$. Consider the "multiple comparisons with a control group" problem, i. e., the hypotheses $H_i : \mu_i = \mu_k$, $1 \leq i \leq k-1$, leading to $m = k-1$. Equivalently, we can express H_i as $\theta_i = 0$, where $\theta_i = \mu_i - \mu_k$ is a derived parameter. In a compact matrix notation, we can express $\mathcal{H}_{k-1} = (H_1, \ldots, H_{k-1})$ as $C_{\text{Dunnett}}\mu = 0$. Line i of the latter system of equations is equal to H_i, $1 \leq i \leq k-1$. The contrast matrix C_{Dunnett} is Dunnett's contrast matrix with $k-1$ rows and k columns, where in each row j the j-th entry equals $+1$, the k-th entry equals -1 and all other entries are equal to zero. This is a classical multiple test problem which has been considered in the pioneering works of Charles W. Dunnett, cf. [16], [17].

Lemma 2: Denoting the empirical mean in group i by $\bar{X}_{i.}$, suitable (standard) test statistics for the two-sided comparisons as defined in Definition 2 are given by $|T_i|$, $1 \leq i \leq k-1$, where $T_i = \sqrt{n_i n_k/(n_i+n_k)}(\bar{X}_{i.} - \bar{X}_{k.})$. The joint distribution of $T = (T_1, \ldots, T_{k-1})^\top$ is multivariate normal (or multivariate t) with a covariance matrix Σ which only depends on the sample sizes $n_1, \ldots, n_k$. More specifically, we have that $T \sim \mathcal{N}_{k-1}(\tilde{\mu}, \Sigma)$ with

$$\tilde{\mu}_i = \sqrt{\frac{n_i n_k}{n_i+n_k}}(\mu_i - \mu_k) \quad \text{and} \quad \Sigma = DC_{\text{Dunnett}}MC_{\text{Dunnett}}^\top D,$$

where $D = \text{diag}\left(\sqrt{\frac{n_i n_k}{n_i+n_k}} : 1 \leq i \leq k-1\right) \in \mathbb{R}^{k-1\times k-1}$ and $M = \text{diag}(n_i^{-1} : 1 \leq i \leq k) \in \mathbb{R}^{k\times k}$.

Proof: The proof is a straightforward application of the linearity of Gaussian distributions and the assertion follows, for instance, immediately from Section 3 in [18]. ∎

Theorem 3: Under the assumptions of Lemma 2, the structural properties (S1) - (S3) are fulfilled for the STP induced by T.

Proof: It remains to show (S1). To this end, notice that for any $\mu^* \in H_0$ the joint distribution of T is identical, namely, a centered Gaussian distribution with covariance matrix Σ. Therefore, the FWER for the induced STP is invariant with respect to $\mu^* \in H_0$. Now, consider $\mu \notin H_0$, with corresponding index set $I_0(\mu)$ of true hypotheses in $\mathcal{H}_{k-1}$. Without loss of generality, assume that $I_0(\mu) = \{1, \ldots, m_0\}$, with $m_0 = |I_0(\mu)|$ denoting the number of true hypotheses in $\mathcal{H}_{k-1}$ under μ. The subvector $(T_1, \ldots, T_{m_0})$ has the same joint distribution under μ and under any $\mu^* \in H_0$, namely, an m_0-dimensional, centered Gaussian distribution with covariance matrix given by the appropriate submatrix of Σ. We conclude that

$$\begin{aligned}\text{FWER}_\mu(\varphi) &= \mathbb{P}_\mu\left(\bigcup_{i=1}^{m_0}\{\varphi_i = 1\}\right) \\ &= \mathbb{P}_{\mu^*}\left(\bigcup_{i=1}^{m_0}\{\varphi_i = 1\}\right) \\ &\leq \mathbb{P}_{\mu^*}\left(\bigcup_{i=1}^{m}\{\varphi_i = 1\}\right) \\ &= \text{FWER}_{\mu^*}(\varphi),\end{aligned}$$

completing the proof. ∎

We may remark here that the latter calculation is an instance where the much more general concept of "subset pivotality" (introduced and extensively been made use of for resampling in [6]) applies.

For ease of graphical illustration, let us now consider the case of $k = 3$ and, consequently, $m = 2$.

Corollary 1: If, under the assumptions of Lemma 2, $k = 3$, we obtain $T \sim \mathcal{N}_2(\tilde{\mu}, \Sigma)$, where $\tilde{\mu}$ is as in Lemma 2 and

$$\Sigma = \begin{pmatrix} 1 & \sqrt{\frac{n_1 n_2}{(n_1+n_3)(n_2+n_3)}} \\ \sqrt{\frac{n_1 n_2}{(n_1+n_3)(n_2+n_3)}} & 1 \end{pmatrix}.$$

Since the joint distribution of T under the global hypothesis is exactly known here, the copula C_{μ^*} of the distributional transforms under H_0 can simply be calculated by transformation of measures. For $(u_1, u_2) \in [0,1]^2$, we obtain $C_{\mu^*}(u_1, u_2) = F_{|T|}(\Phi^{-1}((u_1+1)/2), \Phi^{-1}((u_2+1)/2))$, where Φ denotes the cdf. of the standard normal distribution and $F_{|T|}$ the joint cdf. of the absolute values of T under μ^*, which is easily evaluable by numerical routines for

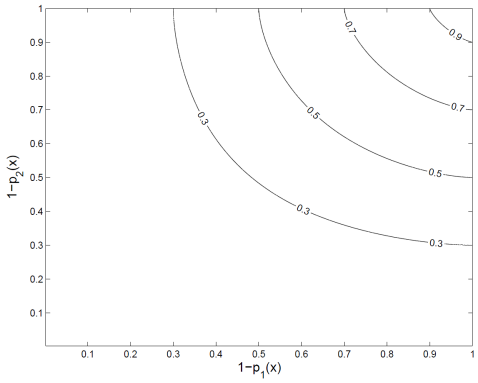

Figure 1. Contour lines of C_{μ^*} in the case of $(n_1, n_2, n_3) = (5, 100, 5)$ for the STP defined by Definition 2 and Lemma 2.

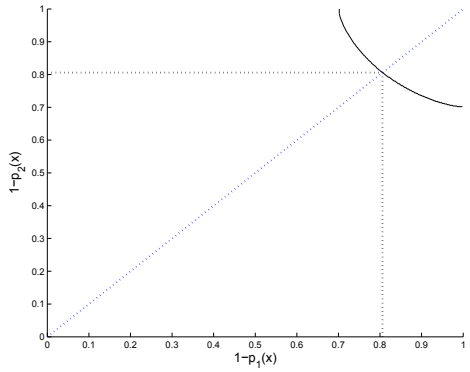

Figure 2. Graphical representation of the construction of an STP according to Theorem 2.

multivariate normal distributions like the `mvtnorm` package in `R`, cf. [19].

Figure 1 depicts contour lines of C_{μ^*} in the case of $(n_1, n_2, n_3) = (5, 100, 5)$ for contour levels $0.3, 0.5, 0.7$, and 0.9. The maybe unrealistic sample sizes were chosen for ease of graphical representation. For control of the FWER at level $\alpha = 0.3$ (say) with the STP defined by Definition 2 and Lemma 2, Figure 2 represents our findings from Theorem 2 graphically: An STP is constructed by determining the point of intersection of the diagonal on $[0, 1]^2$ with the contour line of C_{μ^*} at contour level $1 - \alpha$. Projection onto the coordinate axes yields the multiplicity-adjusted marginal copula arguments $1 - \alpha_{\text{loc.}}^{(i)}, i = 1, 2$. In the example, one could consequently choose local significance levels $\alpha_{\text{loc.}}^{(1)} = \alpha_{\text{loc.}}^{(2)} = 0.1943$.

Remark 1: Since every bivariate $(1 - \alpha)$-quantile of C_{ϑ^*} is a solution to the problem of STP construction according to Theorem 2, Figure 2 furthermore shows how an importance weighting of the individual hypotheses can be incorporated straightforwardly into our method: the only thing that has to be changed is the slope of the line through the origin.

B. Parametric copula estimation

The example in the previous section was mainly intended to provide a proof of principle for the applicability of Theorem 2 and thus, we chose an easy-to-follow setup. However, from a practical point of view this example is not convincing, because the full joint distribution of T under ϑ^* was exactly derivable and therefore, the detour via the distributional transforms is not needed in practice for such type of examples. Much more interesting are cases where prior information about C_{ϑ^*} is incomplete or even lacking. In such cases, two inferential problems have to be solved in parallel, namely (i) estimation of the dependency structure, (ii) multiple testing. From methods based on empirical covariance matrices, it is well-known that the "curse of dimensionality" ($m(m-1)/2$ parameters have to be estimated) requires some kind of regularization, i. e., structural assumptions regarding the data-generating process. Two respective techniques are shrinking of the empirical covariance matrix $\hat{\Sigma}$ (say) toward some pre-specified target (cf., for instance, [20]) or low-rank approximations of $\hat{\Sigma}$ by, for example, assuming equi-correlation in blocks or AR(1) or Toeplitz structures, cf. [21] for applications in the context of multiple testing.

Here, we propose a different approach based on parametric families of copulae and the "realized copula" concept from [22]. Parametric copula families are intensively studied in the textbook [23] and in recent literature from the fields of quantitative finance and econometrics, see, among many others, [24] and [25] for overviews. To illustrate the proposed method, let us consider a real-life dataset from cancer research which can be downloaded freely from the Gene Expression Omnibus data repository, see http://www.ncbi.nlm.nih.gov/sites/GDSbrowser, dataset GDS2771. More detailed information about the underlying studies is given in [26] and [27]. We focus on a specific aspect of the dataset, namely, the determination of genes that are (on average) differentially expressed in airway epithelial cells of cancer patients in comparison with healthy controls. To this end, we restrict attention to $m = 11$ genes on chromosome 1. In dataset GDS2771, expression profiles of $n_1 = 97$ patients and $n_2 = 90$ controls for these m genes are tabulated. In this, the raw expression counts were transformed to marginally fit normal distributions well. Indeed, diagnostic plots (not shown here) confirm that, marginally, Gaussian distributions are valid models. Consequently, our parameter of interest $\vartheta = (\vartheta_1, \ldots, \vartheta_m)^\top$ consists of the differences in mean expression levels of the $m = 11$ genes between the patient group and the control group on the corresponding transformed scales, and $\vartheta^* = \mathbf{0} \in \mathbb{R}^m$. However, how the aforementioned (gene-specific) marginal transformations affect the dependency structure of p-values originating from marginal two-sample t-tests is not at all clear. Therefore, we chose to separate the dependency structure assessment

completely from the marginal models (which is possible by our copula-based STP approach) and considered the flexible class of m-variate Clayton copulae (see, for instance, Example 4.23 in [23]) for the dependency modeling. Each member of the family of Clayton copulae is uniquely defined by a one-dimensional parameter $\eta > 0$ and has the form

$$C_\eta(u_1, \ldots, u_m) = \left(u_1^{-\eta} + u_2^{-\eta} + \ldots + u_m^{-\eta} - m + 1\right)^{-1/\eta}. \quad (3)$$

Before discussing the results for dataset GDS2771, we describe the procedure for empirical calibration of η according to the realized copula method. To this end, the following "inversion formulas" are helpful.

Lemma 3: Let X and Y two real-valued random variables with marginal cdfs F_X and F_Y and bivariate copula C_η, depending on a copula parameter η. Let $\sigma_{X,Y}$, $\rho_{X,Y}$ and $\tau_{X,Y}$ denote (the population versions of) the covariance, Spearman's rank correlation coefficient and Kendall's tau, respectively, of X and Y. Then it holds:

$$\sigma_{X,Y} = f_1(\eta) = \int_{\mathbb{R}^2} [C_\eta\{F_X(x), F_Y(y)\} - F_X(x)F_Y(y)]\, dx\, dy, \quad (4)$$

$$\rho_{X,Y} = f_2(\eta) = 12 \int_{[0,1]^2} C_\eta(u, v)\, du\, dv - 3, \quad (5)$$

$$\tau_{X,Y} = f_3(\eta) = 4 \int_{[0,1]^2} C_\eta(u, v)\, dC_\eta(u, v) - 1. \quad (6)$$

Proof: Equation (4) is due to Hoeffding, see [28], equation (5) is Theorem 5.1.6. in [23] and (6) is Theorem 5.1.3 in [23]. ∎

The "realized copula" method for empirical calibration of a one-dimensional parameter η of an m-variate copula essentially considers every of the $m(m-1)/2$ pairs of the m underlying random variables, inverts (4) each time with respect to η, replaces the population covariance by its empirical counterpart and aggregates the resulting $m(m-1)/2$ estimates in an appropriate way. More specifically, the authors of [22] define for $1 \leq i < j \leq m$: $g_{ij}(\eta) = \hat{\sigma}_{ij} - f_1(\eta)$, set $\mathbf{g}(\eta) = (g_{ij}(\eta))_{1 \leq i < j \leq m}$, and propose to estimate

$$\hat{\eta} = \arg\min_\eta \mathbf{g}^\top(\eta) \mathbf{W} \mathbf{g}(\eta)$$

for an appropriate weight matrix $\mathbf{W} \in \mathbb{R}^{\binom{m}{2} \times \binom{m}{2}}$. In this, $\hat{\sigma}_{ij}$ denotes the empirical covariance of X_i and X_j.

Remark 2: In the realized copula method, any of the functions f_ℓ, $\ell = 1, 2, 3$ corresponding to relationships (4) - (6) may be employed. Moreover, they may be combined to estimate two- or three-dimensional copula parameters η.

Returning to our real-data example, the X_i have to be replaced by the distributional transforms $1 - p_i$ of the t-statistics in each marginal. In order to assess their correlation structure under ϑ^*, we employed a resampling strategy. For a fixed number $B = 1{,}000$, we permuted the entire data vectors of the $n = n_1 + n_2 = 187$ study participants, i. e., we randomly assigned each study participant's data vector to the "cancer positive" or the "cancer negative" group in each permutation run. This resampling mechanism destroys information about the differential expression between the two groups in every marginal (thus reflecting the situation under ϑ^*), but preserves the dependency structure between genes. Similar resampling schemes are made use of extensively for estimating joint distributions of test statistics for multiple test problems from the field of genetics in [7]. After completion of all B permutations and re-calculation of the t-statistics in each permutation run, we utilized the empirical covariances of the resulting resampled distributional transforms as estimates $\hat{\sigma}_{ij}$ in the realized copula optimization step.

Based on this, application of the realized copula method to dataset GDS2771 with η taken as the Clayton copula parameter given in (3) resulted in $\hat{\eta} = 0.1636$, where we treated each gene equally, meaning that we set $\mathbf{W} = I_{\binom{11}{2}}$. Having estimated C_{ϑ^*} in this way, the reasoning of Theorem 2 led, for a target FWER level of $\alpha = 0.05$, to $\alpha_{\text{loc.}}^{(i)} \equiv \alpha_{\text{loc.}} = 0.00467$, $1 \leq i \leq m = 11$. In summary, the empirical calibration of $\alpha_{\text{loc.}}$ based on the intrinsic correlation structure in the data allowed us to enlarge the multiplicity-adjusted local significance level in comparison with the Bonferroni correction (valid under any kind of dependence) and in comparison with the Šidák correction (valid under joint independence of all m marginal tests).

IV. Discussion

We have presented a flexible method to construct simultaneous test procedures based on copulae of distributional transforms, assuming uniformly distributed p-values for every marginal test problem. Useful features of the proposed construction method for STPs are that marginal models and dependency structure can be treated separately both in the modeling step and in the data analysis step of the multiple test problem and that an importance weighting with respect to the individual test problems is straightforwardly possible by choosing an appropriate point on the contour line of the copula of distributional transforms under the global hypothesis. Applicability of the method in practice has been shown in a stylized example case in Section III-A and for a real-life data set from cancer research in Section III-B.

The main limitation of our procedure is that strict FWER control can not be guaranteed for finite sample sizes if the copula parameter is unknown. For large n, consistency of moment estimators, together with Lemma 3, yields approximate FWER control of an STP constructed according to Theorem 2. Future research shall be concerned with the obvious question how the random fluctuations in the data, that lead to noisy estimation of η, translate into fluctuations of the realized FWER of φ around its target level α, and with conservative modifications of our procedure guaranteeing that the FWER of φ can not exceed α, even for noisy $\hat{\eta}$.

REFERENCES

[1] K. R. Gabriel, "Simultaneous test procedures - some theory of multiple comparisons." *Ann. Math. Stat.*, vol. 40, pp. 224–250, 1969.

[2] T. Hothorn, F. Bretz, and P. Westfall, "Simultaneous inference in general parametric models," *Biom. J.*, vol. 50, no. 3, pp. 346–363, Jun 2008.

[3] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *J. R. Stat. Soc., Ser. B*, vol. 57, no. 1, pp. 289–300, 1995.

[4] S. Holm, "A simple sequentially rejective multiple test procedure." *Scand. J. Stat., Theory Appl.*, vol. 6, pp. 65–70, 1979.

[5] A. C. Tamhane, W. Liu, and C. W. Dunnett, "A generalized step-up-down multiple test procedure." *Can. J. Stat.*, vol. 26, no. 2, pp. 353–363, 1998.

[6] P. H. Westfall and S. S. Young, *Resampling-based multiple testing: examples and methods for p-value adjustment.* Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley, New York, 1992.

[7] S. Dudoit and M. J. van der Laan, *Multiple testing procedures with applications to genomics.* Springer Series in Statistics. New York, NY: Springer, 2008.

[8] T. Dickhaus and J. Stange, "Multiple point hypothesis test problems and effective numbers of tests," Sonderforschungsbereich 649, Humboldt-Universität zu Berlin, Germany, SFB 649 Discussion Paper 2012-041, 2012, available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2012-041.pdf.

[9] S. K. Sarkar, "Rejoinder: On methods controlling the false discovery rate," *Sankhyā*, vol. 70, no. 2, Ser. A, pp. 183–185, 2008.

[10] R. Cerqueti, M. Costantini, and C. Lupi, "A copula-based analysis of false discovery rate control under dependence assumptions," Università degli Studi del Molise, Dipartimento di Scienze Economiche, Gestionali e Sociali (SEGeS), Economics & Statistics Discussion Paper 065/12, 2012.

[11] H. Finner and M. Roters, "Multiple hypotheses testing and expected number of type I errors." *Ann. Stat.*, vol. 30, no. 1, pp. 220–238, 2002.

[12] J. D. Storey, J. E. Taylor, and D. Siegmund, "Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach." *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, vol. 66, no. 1, pp. 187–205, 2004.

[13] L. Rüschendorf, "On the distributional transform, Sklar's theorem, and the empirical copula process." *J. Stat. Plann. Inference*, vol. 139, no. 11, pp. 3921–3927, 2009.

[14] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges." *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231, 1959.

[15] ——, "Random variables, distribution functions, and copulas - a personal look backward and forward." in *Distributions with Fixed Marginals and Related Topics.* Institute of Mathematical Statistics, Hayward, CA, 1996, pp. 1–14.

[16] C. W. Dunnett, "A multiple comparison procedure for comparing several treatments with a control." *J. Am. Stat. Assoc.*, vol. 50, pp. 1096–1121, 1955.

[17] ——, "New tables for multiple comparisons with a control." *Biometrics*, vol. 20, pp. 482–491, 1964.

[18] F. Bretz, T. Hothorn, and P. Westfall, "Multiple Comparison Procedures in Linear Models." in *COMPSTAT 2008. Proceedings in Computational Statistics*. Physica Verlag Heidelberg, 2008, pp. 423–431.

[19] A. Genz and F. Bretz, *Computation of multivariate normal and t probabilities.* Lecture Notes in Statistics 195. Berlin: Springer, 2009.

[20] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Stat. Appl. Genet. Mol. Biol.*, vol. 4, Article 32, 2005.

[21] S. Ghosal and A. Roy, "Predicting false discovery proportion under dependence." *J. Am. Stat. Assoc.*, vol. 106, no. 495, pp. 1208–1218, 2011.

[22] M. R. Fengler and O. Okhrin, "Realized Copula," Sonderforschungsbereich 649, Humboldt-Universität zu Berlin, Germany, SFB 649 Discussion Paper 2012-034, 2012, available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2012-034.pdf.

[23] R. B. Nelsen, *An introduction to copulas. 2nd ed.* Springer Series in Statistics. New York, NY: Springer., 2006.

[24] C. Genest, M. Gendron, and M. Bourdeau-Brien, "The Advent of Copulas in Finance." *The European Journal of Finance*, vol. 15, no. 7-8, pp. 609–618, 2009.

[25] W. K. Härdle and O. Okhrin, "De copulis non est disputandum - Copulae: an overview." *AStA Adv. Stat. Anal.*, vol. 94, no. 1, pp. 1–31, 2010.

[26] A. Spira, J. E. Beane, V. Shah, K. Steiling, G. Liu, F. Schembri, S. Gilman, Y. M. Dumas, P. Calner, P. Sebastiani, S. Sridhar, J. Beamis, C. Lamb, T. Anderson, N. Gerry, J. Keane, M. E. Lenburg, and J. S. Brody, "Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer," *Nat. Med.*, vol. 13, no. 3, pp. 361–366, Mar 2007.

[27] A. M. Gustafson, R. Soldi, C. Anderlind, M. B. Scholand, J. Qian, X. Zhang, K. Cooper, D. Walker, A. McWilliams, G. Liu, E. Szabo, J. Brody, P. P. Massion, M. E. Lenburg, S. Lam, A. H. Bild, and A. Spira, "Airway PI3K pathway activation is an early and reversible event in lung cancer development," *Sci. Transl. Med.*, vol. 2, no. 26, p. 26ra25, Apr 2010.

[28] W. Höffding, "Maßstabinvariante Korrelationstheorie." *Schr. math. Inst. Inst. angew. Math. Univ. Berlin*, vol. 5, pp. 181–233, 1940.

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "HMM in dynamic HAC models" by Wolfgang Karl Härdle, Ostap Okhrin and Weining Wang, January 2012.
002 "Dynamic Activity Analysis Model Based Win-Win Development Forecasting Under the Environmental Regulation in China" by Shiyi Chen and Wolfgang Karl Härdle, January 2012.
003 "A Donsker Theorem for Lévy Measures" by Richard Nickl and Markus Reiß, January 2012.
004 "Computational Statistics (Journal)" by Wolfgang Karl Härdle, Yuichi Mori and Jürgen Symanzik, January 2012.
005 "Implementing quotas in university admissions: An experimental analysis" by Sebastian Braun, Nadja Dwenger, Dorothea Kübler and Alexander Westkamp, January 2012.
006 "Quantile Regression in Risk Calibration" by Shih-Kang Chao, Wolfgang Karl Härdle and Weining Wang, January 2012.
007 "Total Work and Gender: Facts and Possible Explanations" by Michael Burda, Daniel S. Hamermesh and Philippe Weil, February 2012.
008 "Does Basel II Pillar 3 Risk Exposure Data help to Identify Risky Banks?" by Ralf Sabiwalsky, February 2012.
009 "Comparability Effects of Mandatory IFRS Adoption" by Stefano Cascino and Joachim Gassen, February 2012.
010 "Fair Value Reclassifications of Financial Assets during the Financial Crisis" by Jannis Bischof, Ulf Brüggemann and Holger Daske, February 2012.
011 "Intended and unintended consequences of mandatory IFRS adoption: A review of extant evidence and suggestions for future research" by Ulf Brüggemann, Jörg-Markus Hitz and Thorsten Sellhorn, February 2012.
012 "Confidence sets in nonparametric calibration of exponential Lévy models" by Jakob Söhl, February 2012.
013 "The Polarization of Employment in German Local Labor Markets" by Charlotte Senftleben and Hanna Wielandt, February 2012.
014 "On the Dark Side of the Market: Identifying and Analyzing Hidden Order Placements" by Nikolaus Hautsch and Ruihong Huang, February 2012.
015 "Existence and Uniqueness of Perturbation Solutions to DSGE Models" by Hong Lan and Alexander Meyer-Gohde, February 2012.
016 "Nonparametric adaptive estimation of linear functionals for low frequency observed Lévy processes" by Johanna Kappus, February 2012.
017 "Option calibration of exponential Lévy models: Implementation and empirical results" by Jakob Söhl und Mathias Trabs, February 2012.
018 "Managerial Overconfidence and Corporate Risk Management" by Tim R. Adam, Chitru S. Fernando and Evgenia Golubeva, February 2012.
019 "Why Do Firms Engage in Selective Hedging?" by Tim R. Adam, Chitru S. Fernando and Jesus M. Salas, February 2012.
020 "A Slab in the Face: Building Quality and Neighborhood Effects" by Rainer Schulz and Martin Wersing, February 2012.
021 "A Strategy Perspective on the Performance Relevance of the CFO" by Andreas Venus and Andreas Engelen, February 2012.
022 "Assessing the Anchoring of Inflation Expectations" by Till Strohsal and Lars Winkelmann, February 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Hidden Liquidity: Determinants and Impact" by Gökhan Cebiroglu and Ulrich Horst, March 2012.
024 "Bye Bye, G.I. - The Impact of the U.S. Military Drawdown on Local German Labor Markets" by Jan Peter aus dem Moore and Alexandra Spitz-Oener, March 2012.
025 "Is socially responsible investing just screening? Evidence from mutual funds" by Markus Hirschberger, Ralph E. Steuer, Sebastian Utz and Maximilian Wimmer, March 2012.
026 "Explaining regional unemployment differences in Germany: a spatial panel data analysis" by Franziska Lottmann, March 2012.
027 "Forecast based Pricing of Weather Derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Matthias Ritter, March 2012.
028 "Does umbrella branding really work? Investigating cross-category brand loyalty" by Nadja Silberhorn and Lutz Hildebrandt, April 2012.
029 "Statistical Modelling of Temperature Risk" by Zografia Anastasiadou, and Brenda López-Cabrera, April 2012.
030 "Support Vector Machines with Evolutionary Feature Selection for Default Prediction" by Wolfgang Karl Härdle, Dedy Dwi Prastyo and Christian Hafner, April 2012.
031 "Local Adaptive Multiplicative Error Models for High-Frequency Forecasts" by Wolfgang Karl Härdle, Nikolaus Hautsch and Andrija Mihoci, April 2012.
032 "Copula Dynamics in CDOs." by Barbara Choroś-Tomczyk, Wolfgang Karl Härdle and Ludger Overbeck, May 2012.
033 "Simultaneous Statistical Inference in Dynamic Factor Models" by Thorsten Dickhaus, May 2012.
034 "Realized Copula" by Matthias R. Fengler and Ostap Okhrin, Mai 2012.
035 "Correlated Trades and Herd Behavior in the Stock Market" by Simon Jurkatis, Stephanie Kremer and Dieter Nautz, May 2012
036 "Hierarchical Archimedean Copulae: The HAC Package" by Ostap Okhrin and Alexander Ristig, May 2012.
037 "Do Japanese Stock Prices Reflect Macro Fundamentals?" by Wenjuan Chen and Anton Velinov, May 2012.
038 "The Aging Investor: Insights from Neuroeconomics" by Peter N. C. Mohr and Hauke R. Heekeren, May 2012.
039 "Volatility of price indices for heterogeneous goods" by Fabian Y.R.P. Bocart and Christian M. Hafner, May 2012.
040 "Location, location, location: Extracting location value from house prices" by Jens Kolbe, Rainer Schulz, Martin Wersing and Axel Werwatz, May 2012.
041 "Multiple point hypothesis test problems and effective numbers of tests" by Thorsten Dickhaus and Jens Stange, June 2012
042 "Generated Covariates in Nonparametric Estimation: A Short Review." by Enno Mammen, Christoph Rothe, and Melanie Schienle, June 2012.
043 "The Signal of Volatility" by Till Strohsal and Enzo Weber, June 2012.
044 "Copula-Based Dynamic Conditional Correlation Multiplicative Error Processes" by Taras Bodnar and Nikolaus Hautsch, July 2012

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

045 "Additive Models: Extensions and Related Models." by Enno Mammen, Byeong U. Park and Melanie Schienle, July 2012.
046 "A uniform central limit theorem and efficiency for deconvolution estimators" by Jakob Söhl and Mathias Trabs, July 2012
047 "Nonparametric Kernel Density Estimation Near the Boundary" by Peter Malec and Melanie Schienle, August 2012
048 "Yield Curve Modeling and Forecasting using Semiparametric Factor Dynamics" by Wolfgang Karl Härdle and Piotr Majer, August 2012
049 "Simultaneous test procedures in terms of p-value copulae" by Thorsten Dickhaus and Jakob Gierl, August 2012

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".